

**Principal Components Analysis (PCA)**

Step 1: Get the data. In this small example we will have  $n = 6$  data points and  $p = 2$  features. In reality we would have many more of each, and sometimes  $p \gg n$ . The data matrix with  $n$  rows and  $p$  columns is called  $X_{\text{orig}}$ :

$$X_{\text{orig}} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad X = \begin{bmatrix} & \\ & \\ & \\ & \\ & \\ & \end{bmatrix}$$

Step 2: Subtract off the column-wise mean from each column (feature) to obtain  $X$  (fill in above). The mean of column  $f$  is:

$$\bar{f} = \frac{1}{n} \sum_{i=1}^n f_i$$

Step 3: Compute the covariance of each pair of features in  $X$  to obtain the  $p \times p$  covariance matrix  $A$ . The covariance of feature  $f$  with feature  $g$  is:

$$\text{cov}(f, g) = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{f})(g_i - \bar{g})$$

Note that in our case, we have set all the means to be 0. Also note that variance is a special case when  $f = g$ :

$$\text{cov}(f, f) = \text{var}(f) = \frac{1}{n-1} \sum_{i=1}^n (f_i - \bar{f})^2$$

Fill in  $A$  below:

$$A = \begin{bmatrix} & \\ & \end{bmatrix}$$

Step 4: Compute the eigenvalues ( $\lambda_1, \lambda_2$  for  $p = 2$ ) and eigenvectors ( $\vec{v}_1, \vec{v}_2$ ) of  $A$ . The eigenvectors (sorted by eigenvalue) will become the directions of our principal components (i.e. new coordinate system). We want our eigenvectors and eigenvalues to satisfy:

$$A\vec{v} = \lambda\vec{v} \quad \Rightarrow \quad \det(A - \lambda I) = 0$$

If you've had linear algebra, verify that the eigenvalues are  $\lambda_1 = \frac{3}{5}$  and  $\lambda_2 = 0$ , and the eigenvectors are  $\vec{v}_1 = [1, -1]^T$  and  $\vec{v}_2 = [1, 1]^T$ . Otherwise use these directly in Step 5.

Step 5: Transform the data  $X$  using the eigenvector matrix  $W$  (one eigenvector on each *column*, sorted by eigenvalue). The number of eigenvectors we use corresponds to the number of dimensions we retain. Say we want to retain  $r$  dimensions, then we would obtain the transformed data  $T_r = XW_r$ .  $T_r$  will be an  $n \times r$  matrix. In our case, use  $r = 2$  and compute  $T_r$ .

$$T_2 = XW_2 = \begin{bmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{bmatrix} \begin{bmatrix} \phantom{0} \\ \phantom{0} \end{bmatrix} = \begin{bmatrix} \phantom{0} \\ \phantom{0} \\ \phantom{0} \\ \phantom{0} \end{bmatrix}$$

Step 6: Finally, plot the transformed data  $T_r$  with principal component 1 (PC1) on the  $x$ -axis and PC2 on the  $y$ -axis. We could plot further PCs on different coordinate systems when  $p > 2$ .